A New Cascade Model for the Hierarchical Joint Classification of Multitemporal and Multiresolution Remote Sensing Data

Ihsen HEDHLI *Student Member, IEEE*, Gabriele MOSER, *Senior Member, IEEE*,
Josiane ZERUBIA, *Fellow, IEEE*, and Sebastiano B. SERPICO, *Fellow, IEEE*

***Abstract*— In this paper, we propose a novel method for the joint classification of both multidate and multiresolution remote sensing imagery, which represents an important and relatively unexplored classification problem. The proposed classifier is based on an explicit hierarchical graph-based model that is sufficiently flexible to address co-registered time series of images collected at different spatial resolutions.**

Within this framework, a novel element of the proposed approach is the use of multiple quad-trees in cascade, each associated with the images available at each observation date in the considered time series. For each date, the input images are inserted in a hierarchical structure on the basis of their resolutions, whereas missing levels are filled in with wavelet transforms of the images embedded in finer-resolution levels. This approach is aimed at both exploiting multiscale information, which is known to play a crucial role in high resolution image analysis, and supporting input images acquired at different resolutions in the input time series. The experimental results are shown for multitemporal and multiresolution optical data.

***Index Terms*— Satellite image time series, multitemporal classification, hierarchical multiresolution Markov random fields.**

I. INTRODUCTION

The capabilities to monitor the Earth's surface, notably in urban and built-up areas, for environmental disasters such as floods or earthquakes and to assess the ground effects and damage of such events play important roles in multiple social, economic, and human viewpoints. In this framework, accurate and time-efficient classification methods are important tools required to support the rapid and reliable assessment of ground changes and damages induced by a disaster, in particular when an extensive area has been affected. Given the substantial amount and variety of data available currently from last-generation very-high resolution (VHR) satellite missions such as Pléiades, COSMO-SkyMed, or WorldView-2 and -3, the main difficulty is to develop a classifier that utilizes the benefits of multiband, multiresolution, multidate, and possibly multisensor input imagery. On the one hand, the use of multiresolution and multiband imagery has been previously shown to optimize the classification results in terms of accuracy and computation time [1] and on the other hand, the integration of the temporal dimension into a classification scheme can significantly enhance the results in terms of accuracy and reliability [2]. Within this framework, classification methods are required that automatically merge the information provided by different sets of images taken on the identical area at different times and resolutions.

In this context, Markov random field (MRF) models are widely used to solve low level processing problems in computer vision and image classification since they provide a convenient and consistent way of integrating contextual information into the classification scheme [3-9]. Because of their generally non-causal nature, MRF models for the spatial context in image classification lead to iterative inference algorithms that are computationally demanding. A well-known example is the optimization via simulated annealing, which is based on the Metropolis-Hastings algorithm, converges under suitable assumption to a globally optimum solution, but typically requires very long times [10-13]. Thus, suboptimal algorithms are often used in practice. For instance, the iterated conditional mode (ICM) is akin to a gradient descent or a simulated annealing frozen at zero temperature. It dramatically reduces computational burden, as compared to simulated annealing, but allows only a locally optimum solution to be reached and may be critically sensitive to initialization. One could also use the modified Metropolis dynamics (MMD), which usually offers a good compromise between computation time and accuracy. In other words, it is often a good tradeoff between a deterministic gradient-like algorithm and the simulated annealing algorithm. More recently, fast methods based on graph theory (e.g., graph cuts) have also become popular. In the

case of binary classification, they allow the globally optimum solution of the maximum *a posteriori* (MAP) criterion to be determined very efficiently. When applied to multiclass problems, they usually reach local maxima with strong optimality properties [43].

By contrast, MRF models defined according to a hierarchical structure exhibit good methodological and application-oriented properties including the following: (i) causality in scale under the Markovianity assumption, allowing the use of non-iterative algorithms with acceptable computational time [14], and (ii) the possibility to incorporate images acquired at multiple resolutions in the hierarchy for multisensor [15] and multiresolution fusion purposes [16, 17].

In the proposed method, the multidate and multiresolution image classification is based on explicit statistical modeling through a hierarchical MRF. The proposed model allows both the input data collected at multiple resolutions and additional multiscale features derived through wavelets to be fused. The proposed approach consists of a supervised Bayesian classifier that combines (i) a joint class-conditional statistical model for pixelwise information and (ii) a hierarchical MRF for spatio-temporal and multiresolution contextual information. Step (i) addresses the modeling of the statistics of the spectral channels acquired at each resolution and conditioned to each class. Step (ii) consists of integrating this statistical modeling in a hierarchical Markov random field for each date.

Given an input time series of remote sensing images acquired at multiple spatial resolutions, the proposed technique is a multiscale and multitemporal model to fuse the related spatial, temporal, and multiresolution information. Two Bayesian approaches can generally be adopted for joint multitemporal classification. The "cascade" approach (e.g., [18]) classifies each image in the input series on the basis of itself and of the previous images. The "mutual" approach classifies each image on the basis of the previous and the subsequent images in the series (e.g., [2] and [19]). Regarding the relationships between the cascade and mutual schemes for multitemporal classification, cascade structures can be considered as a subset of mutual structures. From a different perspective, we could also consider cascade and mutual structures as being associated with ordered and unordered sets of acquisition times. Indeed, this second perspective has been used for long in the area of multitemporal image classification (e.g., [17, 18]), while cascade and mutual approaches have often been related to different categories of applications. Cascade methods have especially been considered for applications requiring to update a previously available land cover map, which was generated by classifying a past image acquisition, on the basis of a new acquisition. Mutual techniques have been focused mostly on applications in which all acquisitions occurred in the past and could be classified jointly. In particular, to deal with causal models, it is necessary to define an order

over the set of images. Accordingly, the cascade approach is adopted in this paper to preserve the temporal ordering of images.

A novel element of the proposed approach is the use of multiple quad-trees in cascade, each associated with each new available image at different dates, to characterize the temporal correlations associated with distinct images in the input time series. The transition probabilities between the scales and between different dates determine the hierarchical MRF because they formalize the causality of the statistical interactions involved.

Specifically, the proposed method jointly addresses two multisource fusion problems: multiresolution data fusion and multitemporal data fusion. On one hand, previous multiresolution image classification techniques have been based, for example, on wavelet transforms [39], low-level hierarchical multiscale segmentation [38], or multiresolution stochastic image models [15, 1], while they have been mostly focused on single-time data. On the other hand, classical change detection methods generally operate with input single-resolution data, while current approaches to multitemporal analysis with optical imagery generally focus on rather long sequences of single scale images with coarser spatial resolutions (e.g., from a few dozens to a few hundreds of meters) and are not aimed at image classification [46][47]. To our best knowledge, the joint problem of multiresolution and multitemporal fusion has been addressed very scarcely in the literature of remote sensing data classification. Partial exceptions are represented by our previous work in [44], which have used hierarchical multitemporal MRFs, and by [42], in which the problems of multitemporal and multiresolution classification of a VHR satellite image time series (SITS) are jointly addressed by combining multitemporal analysis with conditional random fields (CRFs). In [42], a multitemporal CRF is built on an image grid that, in addition to the spatial neighborhood relations, is expanded by temporal interaction terms that link neighboring epochs via transition probabilities between different classes. Furthermore, in [41] the authors modified slightly the hierarchical technique developed in [1] for change detection purpose. The hierarchical algorithm was applied to a combined image using a normalized change index.

This paper is organized as follows. In Section II, we focus on the proposed multitemporal hierarchical model. Section III explains the methodological selections associated with the modeling of transition probabilities and class-conditional statistics. The experimental results of employing this new hierarchical model in time series classification are presented in Section IV. Finally, we conclude and discuss possible directions for future work in Section V. Proofs are reported in the Appendix.

II. METHODOLOGY

The methodological formulation of the proposed approach to multiresolution and multitemporal fusion for classification purposes is described in this section. Specifically, the proposed hierarchical multidate MRF model and its topology are described in Section II.A. Their use for classification purposes is formulated in a Bayesian framework in Section II.B.1, while the related inference and parameter estimation issues are addressed in Sections II.B.2 and II.B.3.

A. Multitemporal hierarchical MRF model

The objective of this study is to develop a method for multitemporal and multiresolution classification of optical images based on a hierarchical Markovian model. Thus, this method has two requirements for accomplishing this task: (i) the method should be parallel to handle the data in a short time and (ii) the method should provide a structure simplifying the interactions between different images in the input data set.

Parallel multigrid (or pyramidal) schemes are one of the possible approaches satisfying requirement (i). The pyramid structure is a type of signal representation in which images are organized according to their resolutions [20] (see Figure 1), i.e., a pyramid P is a stack of images I_n for which the scale $n \in [0, R]$ and R is the height of the pyramid. An element of this pyramid is called a node (i.e., a pixel or group of pixels in the image domain).

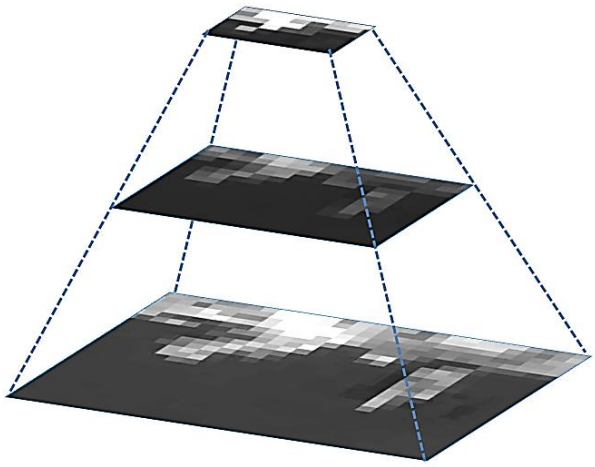


Figure 1: Images are organized according to their resolutions in a pyramid structure

To handle requirement (ii), we define for each node of the pyramid a set of links to other nodes to model scale-to-scale interactions. The theory of multiscale signals has been widely studied, and their representations lead naturally to models of signals on trees. Among others, dyadic trees [21] and quad-trees [1] have been proposed as attractive candidates for modeling these scale-to-scale interactions in mono-dimensional and bi-dimensional signals, respectively. The selection of these structures is justified by their causality properties over scale [22] and possibility of employing a fast optimization method.

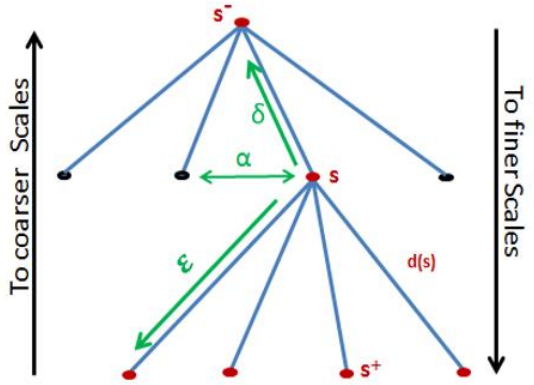


Figure 2: Quad-tree that models the scale-to-scale interactions between images in the pyramid structure. Given a site s in the quad-tree, s^- denotes the parent site of s, s^+ is the set of four children sites of s, $d(s)$ is the set of all descendants of s, and α, δ, and ε are the operators associated with parent, children, and same-scale interchange relationships, respectively.

Let us denote a generic node on the quad-tree as s and the finite set of all nodes as S ($s \in S$). Each node is a pixel in one of the levels of the tree. The set of nodes is hierarchically partitioned, i.e., $S = S^0 \cup S^1 \cup \cdots \cup S^R$ where S^n indicates the subset of nodes associated with the n^{th} level of the tree ($n = 0, 1, 2, \ldots, R$), $n = R$ denotes the root of the tree (coarsest resolution), and $n = 0$ indicates its leaves (finest resolution). In the considered structure, a parent-child relationship can be defined [21]: an upward shift operator δ such that $s^- = \delta(s)$ is the parent of node s. The operator δ is not one-to-one, but four-to-one because each parent has four offsprings (because of the quad-tree structure). We define the forward shift operator ε such that $s^+ = \varepsilon(s)$ is the set of all the descendants of s, the interchange operator α is defined as between the nodes in the identical scale, and $d(s)$ is the set including s and all its descendants in the tree as illustrated in Figure 2. This framework allows data from sensors with different resolutions and different spectral bands to be fused.

A novel element of the proposed approach is the multitemporal aspect. We employ multiple pyramids and quad-trees in a cascade, each pyramid being associated with the set of images available at a different date to characterize the temporal correlations associated with distinct images in the input time series. We build new operators to link between the nodes across different dates. Therefore, we define an upward shift operator ω such that $s^{=} = \omega(s)$ is the parent of node s in the previous date of the time series. Furthermore, we define an interchange operator σ between nodes in the identical scale and identical position but from consecutive dates to characterize the temporal correlation between images given at different dates (Figure 3).

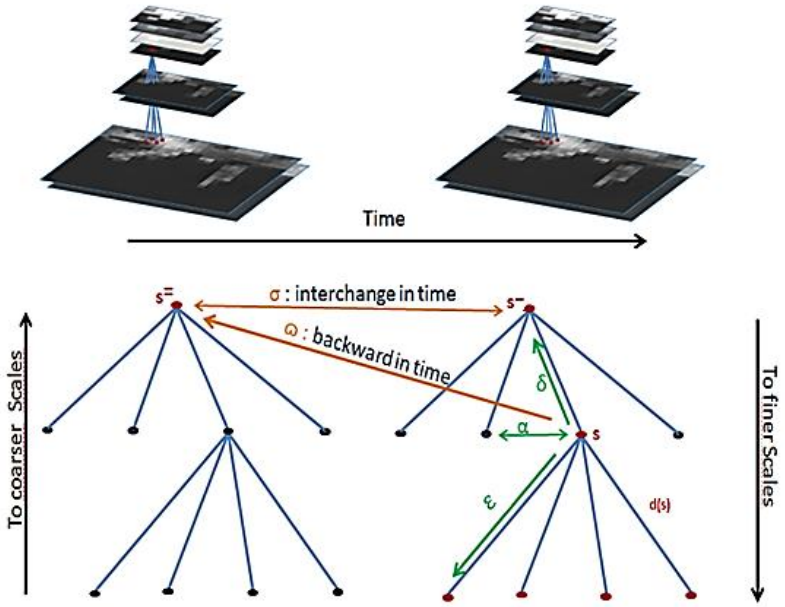


Figure 3: Example of multitemporal hierarchical structure using two quad-trees in cascade. In addition to the same notations used in Fig. 2, $s^{=}$ denotes the parent of site s in the quad-tree corresponding to the previous acquisition time; σ and ω are the operators associated with the relationships between s and $s^{=}$ and between s^{-} and $s^{=}$, respectively.

This multitemporal hierarchical structure is aimed at supporting the joint classification of both multitemporal and multiresolution input images. This implies that, if only one image is available on a certain acquisition time, then it will be included in the finest resolution layer (level 0) of the corresponding quad-tree. Hence, the intrinsic resolution of the image will be the finest resolution of the quad-tree of that date, and all other levels of the tree will be filled in using wavelet transforms of the input image [15].

If images corresponding to multiple resolutions are available on a certain time, a scenario that occurs, for example, when there are both higher resolution panchromatic and coarser resolution multispectral data,

then each single-resolution input image is included in a separate level of the quad-tree. In this case, an assumption implicit in the quad-tree topology is that the spatial resolutions of the input images are related by a power-to-2 relationship. This condition is satisfied with minor approximations by most current multiresolution spaceborne optical sensors (see Section I), so it is operatively only a mild restriction. After the input images are included in the layers corresponding to their spatial resolutions, level 0 of the quad-tree corresponds to the finest-resolution input image, while some intermediate levels generally remain empty and are filled in using wavelet transforms of the images on the lower (finer resolution) layers. For example, if an IKONOS acquisition composed of a panchromatic component at 1-m resolution and a multispectral component at 4-m resolution is used on a certain date, then the finest resolution of the quad-tree is 1 m, the panchromatic and multispectral images are included on levels 0 and 2, respectively, and level 1 is computed as a wavelet transform of the panchromatic data of level 0.

The proposed hierarchical structure allows, in a natural way, the use of an explicit statistical model through a hierarchical Markov random field formulation using a series of random fields at varying scales and times using the operators defined above on the consecutive quad-trees. Let us denote the class label of site s as a discrete random variable x_s and its value as ω_s ($s \in S$). If there are M classes in the considered scene, then each label occupies a value in the set $\Lambda = \{0, 1, \ldots, M-1\}$ (i.e., $\omega_s \in \Lambda$). The class labels of all pixels can be collected in a set $\chi = \{x_s\}_{s \in S}$ of random fields $\chi_t^n = \{x_s\}_{s \in S_t^n}$ associated with each scale n and date t, where S_t^n is the related set of lattice points. The corresponding configuration at scale n and date t can be represented as $\omega_t^n = \{\omega_s\}_{s \in S_t^n}$. The configuration space $\Omega = \Lambda^{|S|}$ is the set of all global discrete labelings (i.e., $\chi \in \Omega$).

We then assume the following to fit an MRF model to the aforementioned hierarchical structure:

– The fundamental assumption of the model is that the sequence of random fields from coarse to fine scales form a Markov chain over scale and time:

$$p\left(\chi_t^n = \omega_t^n \middle| \chi_p^q = \omega_p^q, p < t, q > n\right) = p(\chi_t^n = \omega_t^n | \chi_t^{n+1} = \omega_t^{n+1}, \chi_{t-1}^{n+1} = \omega_{t-1}^{n+1}), \quad (1)$$

– The transition probabilities of this Markov chain factorize so that the components (pixels or nodes) of χ_t^n are mutually independent given χ_t^{n+1} and χ_{t-1}^{n+1} :

$$p(\chi_t^n = \omega_t^n | \chi_t^{n+1} = \omega_t^{n+1}, \chi_{t-1}^{n+1} = \omega_{t-1}^{n+1}) = \prod_{s \in S_t^n} p\left(x_s = \omega_s \middle| x_{s^-} = \omega_{s^-}, x_{s^=} = \omega_{s^=}\right). \quad (2)$$

Using the quad-tree structure allows benefiting from the good properties discussed in Section I (e.g., causality) and applying non-iterative algorithms, thus resulting in a decrease in computational time compared to iterative optimization procedures over graphs.

B. The proposed multitemporal classifier

1) Bayesian Framework

The aim of the classification is to estimate the value $x = \{x_s\}_{s\in S}$ of the hidden label field χ given a realization $y = \{y_s\}_{s\in S}$ of a random field of observations Y attached to the set of nodes S and composed of the input satellite data and of the derived wavelet transforms (see Section II.A). In this context, we consider the problem of inferring the "best" configuration $\hat{x} \in \Omega$. The standard Bayesian formulation of this inference problem consists of minimizing the Bayes risk [23]:

$$\hat{x} = arg \min_{\omega\in\Omega} E\{C(\omega, x)|Y = y\}, \tag{3}$$

where C is the cost function penalizing the discrepancy between the estimated configuration and the "ideal" random configuration and $E\{\cdot\}$ is the expectation operator.

Among the different classification algorithms employed on a quad-tree structure in the literature, two have been widely used. The first algorithm aims to estimate exactly the MAP configuration. The cost function of this algorithm is defined by the following:

$$C_{MAP}(\omega, \omega') = 1 - \delta(\omega, \omega') = 1 - \prod_{s\in S} \delta(\omega_s, \omega'_s), \tag{4}$$

where $\delta(\cdot)$ is the Kronecker delta (i.e., $\delta(a, b) = 1$ for $a = b$ and $\delta(a, b) = 0$ otherwise). This function implies the same cost for all pairs of configurations that differ in, at least, one site. From (3) and (4), the MAP estimator of the label field χ is given by the following:

$$\hat{x}_{MAP} = arg \max_{\omega\in\Omega} p(\chi = \omega | Y = y) \tag{5}$$

This combinatorial optimization problem can be resolved by using a Kalman-like filter, owing to the formal similarity between MRF models and the spatio-temporal models used in Kalman approaches for optical flow [24], or a Viterbi algorithm [25]. The extension of the Viterbi algorithm, which computes the exact MAP estimate of χ given $Y = y$ on the quad-tree has been introduced by Dawid in the context of probabilistic expert systems [45], and Laferté et al. in the context of image classification [1] by proposing a non-iterative algorithm on the quad-tree. However, these algorithms exhibit two main shortcomings. First, computationally, they are known to be affected by underflow problems because of the small

probabilities involved [51]. Second, according to (4), the MAP cost function penalizes the discrepancies between configurations regardless of their corresponding scales, an undesirable property from the viewpoints of segmentation, labeling, and classification [3]. Specifically, an error at a coarser scale will be paid the same cost as an error at a finer scale whereas it is desirable to have a higher cost for errors at coarser levels because they may generally lead to the misclassification of groups of pixels at level 0 (e.g., one pixel at the root corresponds to 4^R pixels at the finest scale).

On the contrary, the marginal posterior mode (MPM) rule is based on a criterion function that aims at segmentation accuracy and allows errors on distinct scales to be penalized differently [1, 3]. The cost function is:

$$C_{MPM}(\omega,\ \omega') = \sum_{s\in S}[1-\delta(\omega_s\ ,\omega'_s)], \tag{6}$$

which is related to the number of sites in which two label configurations differ. The MPM criterion penalizes errors according to their number, consequently to the scale at which they occur. The Bayesian estimator resulting from (6) is given by the following:

$$\forall\ s\in S,\ \hat{x}_s = arg\max_{\omega_s\in\Lambda} p(x_s=\omega_s\,|Y=y), \tag{7}$$

which produces the configuration that maximizes at each site s the a posteriori marginal distribution of x_s conditioned to all observations y. Furthermore, as shown in [48, 49], MPM well adapts the estimator to the quad-tree topology. Indeed, because the tree is acyclic, the labels are estimated recursively by MPM through a forward-backward algorithm similar to the classical Baum and Weiss algorithm for Markov chains [26].

In the following, the explicit distinction between random fields (or variables) and their realizations will be dropped for ease of notation, and for example, the posterior marginal $p\left(x_s=\omega_s\,|Y=y\right)$ will be simply denoted as $p\left(x_s\,|y\right)$.

2) Multitemporal MPM inference

In this study, based on the arguments in Section II.1, an extended version of MPM is developed for the proposed multitemporal hierarchical structure. The posterior marginal $p\left(x_s\,|y\right)$ of the label of each spatio-temporal node s is expressed as a function of the posterior marginal $p\left(x_{s^-}\,|y\right)$ of the parent

node s^- in the corresponding quad-tree and the posterior marginal $p\left(x_{s^=}|y\right)$ of the parent node $s^=$ in the quad-tree associated with the previous date to characterize the temporal correlations associated, at different scales, with distinct images in the input time series. The posterior marginal of (7) can be written as the following:

$$\boldsymbol{p}\left(\boldsymbol{x_s}|\boldsymbol{y}\right) = \sum_{x_{s^-},\, x_{s^=}} \left[\frac{p\left(x_s,\, x_{s^-}, x_{s^=} \mid y_{d(s)}\right)}{\sum_{x_s} p\left(x_s,\, x_{s^-}, x_{s^=} \mid y_{d(s)}\right)} . \boldsymbol{p}\left(\boldsymbol{x_{s^-}}|\,\boldsymbol{y}\right) \boldsymbol{p}\left(\boldsymbol{x_{s^=}}|\boldsymbol{y}\right) \right] \tag{8}$$

where bold font denotes the marginal posteriors of interest to the MPM. This equation involves two conditional independence assumptions:

A1. The label x_s, given the labels of the parents x_{s^-} and $x_{s^=}$ on the same and the previous dates, depends only on the observations $y_{d(s)}$ of site s and of its descendants and not on the observations of the other sites, i.e., $p\left(x_s|y,\, x_{s^-}, x_{s^=}\right) = p\left(x_s \mid y_{d(s)},\, x_{s^-}, x_{s^=}\right)$;

A2. Given the observations, the label of the parent s^- of a site s on the same date is independent on the label of the parent $s^=$ on the previous date, i.e., $p\left(x_{s^-}|x_{s^=}, y\right) = p\left(x_{s^-}|y\right)$.

These assumptions are analogous to the conditional independence assumptions that are commonly accepted when dealing with (hierarchical or single-scale) MRF-based image analysis. They are used within the proposed method for analytical convenience. Proof of (8) can be found in the Appendix.

This formulation allows calculating recursively the posterior marginal $p\left(x_s|y\right)$ at each spatio-temporal node s while the probabilities $p\left(x_s,\, x_{s^-}, x_{s^=} \mid y_{d(s)}\right)$ are produced. Indeed, using arguments similar to [1], the following is proven in the Appendix:

$$p\left(x_s,\, x_{s^-}, x_{s^=} \mid y_{d(s)}\right) = p\left(x_s \mid x_{s^-},\, x_{s^=}\right) . \frac{p\left(x_{s^-} \mid x_{s^=}\right) . p\left(x_{s^=}\right)}{p\left(x_s\right)} . p\left(x_s \mid y_{d(s)}\right), \tag{9}$$

under the following further conditional independence assumption:

A3. The distribution of the labels s^- and $s^=$ of the parents of a site s are independent on the observations $y_{d(s)}$ of the descendants of s, when conditioned to the label x_s of s, i.e., $p(x_{s^-}, x_{s^=}|x_s, y_{d(s)}) = p(x_{s^-}, x_{s^=}|x_s)$.

In (9), the first factor $p\left(x_s \mid x_{s^-},\, x_{s^=}\right)$ corresponds to the child-parent transition probability; $p\left(x_s\right)$ is the prior probability; $p\left(x_{s^-} \mid x_{s^=}\right)$ is the temporal transition probability in the same scale; and $p\left(x_s \mid\, y_{d(s)}\right)$ is the partial posterior marginal probability.

To compute these probabilities, we benefit from the hierarchical structure defined above and use three recursive passes on the quad-tree, including one "bottom-up" and two "top-down" passes (see algorithm 1). For the sake of brevity, only the steps associated with a pair of images acquired on two different times ($t = 0$ and $t = 1$) are explained in the following (see Figure 4). The recursive extension to more than two acquisition times is straightforward.

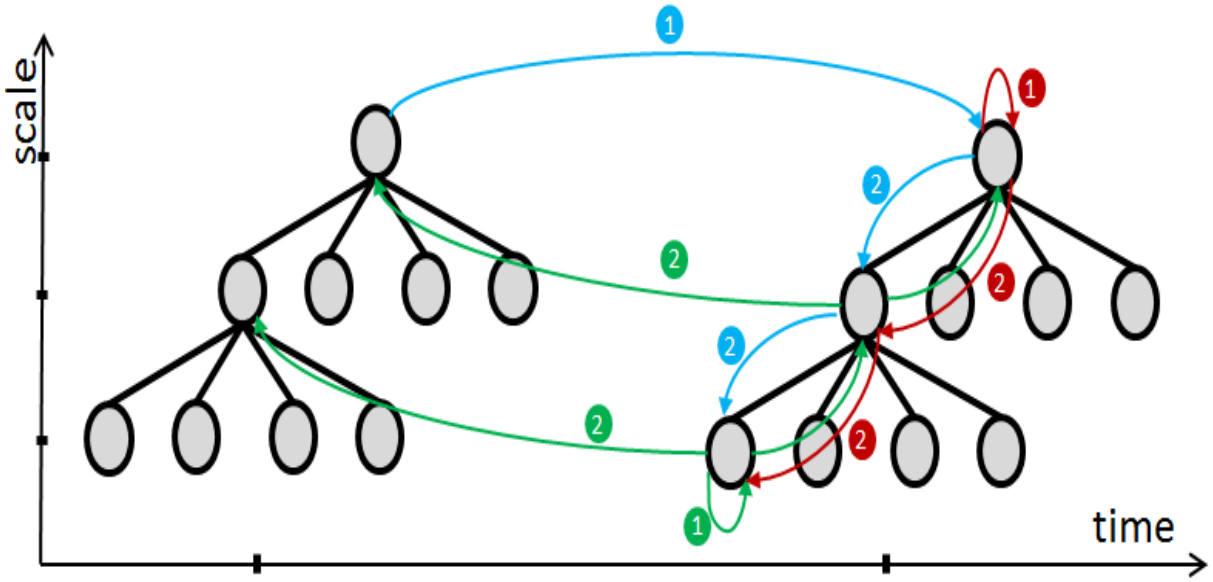


Figure 4: Multitemporal recursive formulation of the MPM criterion on a sequence of two quad-trees (i.e., two acquisition times). The case $R = 2$ is considered as an example. Blue, green, and red arrows indicate the calculations performed by the first top-down, the bottom-up, and the second top-down passes. For each pass, numbers 1 and 2 indicate initialization and recursive computation, respectively.

Algorithm 1. Multitemporal MPM

1. Input: classification map in the root of quad-tree at time $t = 0$; number T of observations date.

For t from 0 to $T - 1$:

2. Initialize: prior initialization via (10)
3. Top down pass: prior estimation via (11)
4. Bottom-up pass: estimation of $p\left(x_s \mid y_{d(s)}\right)$ and $p\left(x_s, x_{s^-}, x_{s^=} \mid y_{d(s)}\right)$ via (9) and (13)
5. Top down pass: estimation of $p\left(x_s \mid y\right)$ at each level of the quad-tree via (8)
6. Output: maximization of $p\left(x_s \mid y\right)$ using *MMD*
7. $t \leftarrow t + 1$.

a) **Time $t = 0$: single-time MPM.** According to the cascade approach, first, classification is performed at time $t = 0$ using a single-date MPM as in [1] and [6], in which the segmentation is obtained recursively over scales through a top-down and a bottom-up stages. Details of this single-date formulation can be found in [1]. We only recall that the process is initialized by predefining the pixelwise prior probability distribution on the root of the corresponding quad-tree, i.e., $p(x_s)$, $s \in S_0^R$. This initialization is required to begin a top-down recursion and compute the priors in all levels of the quad-tree at time 0. A simple initialization strategy is to use a uniform prior distribution on Λ. Here, to incorporate spatial contextual information and mitigate possible blocky artifacts [1, 51], a case-specific initialization strategy is applied that makes use of a spatial MRF model: a neighborhood system is defined on the lattice S_0^R in the root at time 0, and for each pixel $s \in S_0^R$, the unconditional prior $p(x_s)$ is replaced by the local conditional prior $p(x_s|x_{s'}, s' \sim s, s' \in S_0^R)$, where $s \sim s'$ denotes that the sites s and s' are neighbors. This choice generally provides a biased prior-probability estimate but favors spatial adaptivity, a desired property when working with high resolution images in which spatial details are common.

The well-known Potts MRF model, which favors the same labeling in homogeneous image regions, is used [3, 29], i.e.:

$$p\left(x_s \,|x_{s'}\, s' \sim s, s' \in S_0^R\right) \propto exp\left(-\beta \sum_{s \sim s'} \delta\left(x_s, x_{s'}\right)\right), \tag{10}$$

where β is a positive spatial smoothness parameter. Several methods have been proposed to optimize the value of this parameter including the maximization of the pseudo-likelihood function over the training set [28]. In [52], Chardin et al. also combine a hierarchical structure and the Potts model, leading to a semi-

iterative technique in which the Potts component is used to compute a unique prior distribution for each scale. On the contrary, here, we use the Potts model to define a local characteristic for each node of the root level to maximize spatial adaptivity.

As a result of single-time processing at time $t = 0$, the posterior marginal $p(x_s|y)$ is known for each pixel of the corresponding quad-tree; $p(x_s|y_{d(s)})$, in which $y_{d(s)}$ denotes the collection of the observations of all quad-tree sites that are descendants of site s, is also derived as a by-product ($s \in S_0^n, n = 0,1, \dots, R$). Details can be found in [1].

b) **Time $t = 1$: first top-down pass.** In the proposed method, the recursive top-down / bottom-up formulation used for the single-time case in [1] is extended to the multitemporal classification at time 1. In this case as well, first, the prior distribution on the root lattice, i.e., $p(x_s)$, $s \in S_1^R$, has to be defined to initialize a top-down pass. Following the cascade approach, at time 1, we take benefit of the inference conducted at time 0: for each pixel $s \in S_1^R$ on the root lattice at $t = 1$, the unconditional prior $p(x_s)$ is initialized as the posterior marginal $p\left(x_{\sigma(s)}|y_{d[\sigma(s)]}\right)$, which corresponds to the same pixel $\sigma(s) \in S_0^R$ in the root lattice S_0^R at $t = 0$ (blue arrow labeled with the number one in Figure 4) and has been computed as a by-product of the single-date MPM application at time 0.

After initializing the prior in the root, a top-down pass (blue arrow labeled with the number two in Figure 4) is performed for each finer level $n < R$ at time 1. The prior-probability distribution is derived as a function of the prior-probability distribution at the parent level and of the transition probabilities from the parent to the current level ($s \in S_1^n, n = 0,1, \dots, R-1$):

$$p\left(x_s\right) = \sum_{x_{s^-}} p\left(x_s \mid x_{s^-}\right) . p\left(x_{s^-}\right) \tag{11}$$

This derivation favors an identical parent-child labeling and models the statistical interactions between consecutive levels of the quad-tree. We model the transition probability in the form introduced by Bouman et al. [26], i.e., ($s \in S_1^n, n = 0,1 \cdots, R-1$):

$$p\left(x_s \mid x_{s^-}\right) = \begin{cases} \theta & x_s = x_{s^-} \\ \dfrac{1-\theta}{M-1} & x_s \neq x_{s^-} \end{cases}, \tag{12}$$

where θ is a parameter ranging in $[\frac{1}{M}, 1]$. As a result of the first top-down pass, the prior distribution $p(x_s)$ is derived for each pixel s ($s \in S_1^n, n = 0,1, \dots, R$) of each level of the quad-tree at time $t = 1$.

c) Time $t = 1$: bottom-up pass. A bottom-up pass recursion is then performed to estimate the joint probabilities $p\left(x_s\,,\, x_{s^-}\,, x_{s^=} \mid y_{d(s)}\right)$ starting from the leaves of the quad-tree at time 1 and proceeding until the root is reached based on the factorization in (9).

In addition to priors, which have been computed in the previous top-down pass, three sets of probabilities are required to compute this factorization: (i) the set of temporal transition probabilities at the same scale $p\left(x_{s^-} \mid x_{s^=}\right)$; (ii) the child-parent transition probability $p\left(x_s \mid x_{s^-}\,,\, x_{s^=}\right)$; and (iii) the partial posterior marginals $p\left(x_s \mid y_{d(s)}\right)$. Details of the calculation of (i) and (ii) are shown in Section II.B.3. Concerning (iii), Laferté et al. [1] proved the following ($s \in S_1^n, n = 1,2,\cdots,R$):

$$p\left(x_s \mid y_{d(s)}\right) \propto\ p\left(y_s \mid x_s\right) \cdot p\left(x_s\right) \cdot \prod_{\tilde{s} \in s^+} \sum_{x_{\tilde{s}}} \left[\frac{p\left(x_{\tilde{s}} \mid y_{d(\tilde{s})}\right)}{p\left(x_{\tilde{s}}\right)} \cdot p\left(x_{\tilde{s}} \mid x_s\right) \right] \tag{13}$$

Thus, the bottom-up pass is a recursion that estimates $p\left(x_s \mid y_{d(s)}\right)$. It starts from the leaves of the quad-tree in which the partial posterior marginals are computed via (green arrow labeled with the number 1 in Figure 4):

$$p\left(x_s \mid y_s\right) \propto\ p\left(y_s \mid x_s\right) \cdot p\left(x_s\right), \tag{14}$$

and then proceeds until the root is reached using (13) (green arrow labeled with the number two in Figure 4). (13) involves the pixelwise class-conditional PDFs $p\left(y_s \mid x_s\right)$ of the image data at each node of each quad-tree (see Section II.B.3). As a result of the bottom-up pass, we now have all needed probabilities to compute $p\left(x_s\,,\, x_{s^-}\,, x_{s^=} \mid y_{d(s)}\right)$ at each level of the quad-tree.

d) Time $t = 1$: second top-down pass. According to (8), first, the posterior marginal is initialized at the root of time 1 (red arrow labeled with the number one in Figure 4). For this purpose, we initialize $p\left(x_s \mid y\right)$ as $p\left(x_s \mid y_{d(s)}\right)$ for $s \in S_1^R$, as in the usual single-date formulations of MPM [1] Then, the posterior $p\left(x_s \mid y\right)$ at each pixel s for all other tree levels at time $t = 1$ ($s \in S_1^n, n = 0,1,\ldots,R-1$) can be easily computed recursively in a top-down pass (red arrow labeled with the number two in Figure 4) using the formulation in (8).

e) Both times: combination with MMD. At each time $t \in \{0,1\}$, the aforementioned steps lead to the computation of the posterior marginal $p\left(x_s \mid y\right)$ on each pixel ($s \in S_t^n, n = 0,1,\ldots,R$). In principle, the

class label x_s that maximizes $p\left(x_s \mid y\right)$ over the finite set Λ of classes could be selected and assigned to s. This is a feasible procedure but is often avoided in the literature of hierarchical MRFs because of its computational burden (linear with respect to the number of classes and the number of sites in all scales and times) and of possible blocky artifacts [1, 15, 51]. As an alternate approach, here, a case-specific formulation of MMD is applied separately for each scale and time. Specifically, in the case of the root layer of the quad-tree corresponding to each time t, MMD is used to minimize the following energy with respect to the label configuration $\chi_t^R = \{x_s\}_{s \in S_t^R}$:

$$U(\chi_t^R|y) = -\sum_{s \in S_t^R} \ln p(x_s|y) - \beta \sum_{\substack{s,s' \in S_t^R, \\ s \sim s'}} \delta(x_s, x_{s'}) \tag{15}$$

where the first term is expressed in terms of the pixelwise posteriors computed by MPM and the second contribution is due to the Potts model on the root of the tree. MMD is iterative and is initialized with a randomly generated configuration of the label field χ_t^R. At each iteration, it randomly draws one pixel $s \in S_t^R$ and a candidate label for s using a uniform distribution: if this label yields a decrease in $U(\chi_t^R|y)$, then it is assigned to s; otherwise, it is discarded [29].

In the case of each other layer $n = 0,1, \ldots, R-1$, no Potts model is used and MMD is applied to minimize:

$$U(\chi_t^n|y) = -\sum_{s \in S_t^n} \ln p(x_s|y). \tag{16}$$

This means that, in this case, MMD is equivalent to iteratively selecting a random subset of pixels for which random replacements in class membership are attempted. In all cases, the iterative procedure of MMD is repeated until the difference in energy on consecutive iterations goes below a predefined threshold (which was set to 10^{-4} in the experiments).

In the case of the root layer, the solutions obtained using MMD and maximizing $p(x_s|y)$ directly intrinsically differ because the former takes into account spatial context through the Potts model while the latter does not. In the case of the other layers, MMD basically acts as a randomized version of the maximization of $p(x_s|y)$ on every pixel. Computationally, the number of iterations of MMD that suffices to reach convergence is usually significantly smaller than the number of individual operations leading to the maximization of $p(x_s|y)$ on every pixel with respect to the class label. Accordingly, MMD is expected to be advantageous from a computational viewpoint. This is consistent with various previous works using

MPM on hierarchical MRF models (see, e.g., [51] that combines MPM and ICM; and [15] with MPM and MMD).

3) Transition Probabilities and pixelwise class-conditional PDFs

a) Transition probabilities

The transition probabilities between consecutive scales and consecutive dates determine the properties of the hierarchical MRF because they formalize the causality of the statistical interactions involved. Therefore, they must be carefully defined.

In the proposed method, two types of probabilities involve time. The first is the set of temporal transition probabilities at the identical scale $p\left(x_{s^-} \mid x_{s^=}\right)$, which are estimated using a specific formulation of the expectation-maximization (EM) algorithm [34]. An iterative fixed-point EM-like algorithm is performed to estimate the prior joint probabilities $p\left(x_{s^-}, x_{s^=}\right)$ for each scale n, and the temporal transition probabilities are then derived [19]. The probabilities $p\left(x_{s^-} = m, x_{s^=} = m'\right)$, where m and m' range in $\Lambda = \{0,1, \ldots, M-1\}$, are regarded as the elements of an $M \times M$ matrix J, which is computed by maximizing the following pseudo-likelihood ($n = 0,1, \ldots, R$):

$$L(J) = \prod_{s \in S_1^n} \left(\sum_{x_{s^-}} \sum_{x_{s^=}} p\left(x_{s^-}, x_{s^=}\right) p(y_{s^-}, y_{s^=} | x_{s^-}, x_{s^=}) \right). \tag{17}$$

The recursive equation to be used to maximize (17) is the following:

$$p_{k+1}\left(x_{s^-}, x_{s^=}\right) \propto \sum_{s \in S_1^n} \frac{p_k\left(x_{s^-}, x_{s^=}\right) p\left(y_{s^-} \middle| x_{s^-}\right) p\left(y_{s^=} \middle| x_{s^=}\right)}{\sum_{x_{s^-}} \sum_{x_{s^=}} p_k\left(x_{s^-}, x_{s^=}\right) p\left(y_{s^-} \middle| x_{s^-}\right) p\left(y_{s^=} \middle| x_{s^=}\right)}, \tag{18}$$

where $p_k\left(x_{s^-}, x_{s^=}\right)$ is the iterative joint probability estimate at the k^{th} EM iteration. These estimates are initialized by assigning equal probabilities to each pair of classes:

$$p_0\left(x_{s^-}, x_{s^=}\right) = \frac{1}{M^2} \tag{19}$$

The second type of transition probabilities that involve time is the child-parent transition probability $p\left(x_s \mid x_{s^-}, x_{s^=}\right)$. To our knowledge, a case-specific formulation of EM is not available for inter-scale transition probabilities. However, parametrically modeling these probabilities have demonstrated an effective choice in the case of single-date classification as it allowed accurate results to

be obtained [1, 32]. Indeed, we extend here the model proposed by Bouman and Shapiro [32], which favors the identity between the children and parents (in the current and previous dates), all other transitions being unlikely:

$$p\left(x_s \mid x_{s^-}, x_{s^=}\right) = \begin{cases} \theta & x_s = x_{s^-} = x_{s^=} \\ \varphi & (x_s = x_{s^-} \; or \; x_s = x_{s^=}) \; and \; x_{s^-} \neq x_{s^=} \\ \dfrac{1-\theta}{M-1} & x_s \neq x_{s^-} \; and \; x_s \neq x_{s^=} \; and \; x_{s^-} = x_{s^=} \\ \dfrac{1-2\varphi}{M-2} & x_s \neq x_{s^-} \; and \; x_s \neq x_{s^=} \; and \; x_{s^-} \neq x_{s^=} \end{cases} \quad (20)$$

with the parameters $\theta > 1/M$ and $1/M < \varphi < 1/2$. Here, θ has the same meaning as in (12), and the same parameter value is used in both transition probabilities.

b) Pixelwise class-conditional PDFs

Given a training set for each input date, for each class m, scale n and acquisition time t we model the corresponding class-conditional marginal PDF $p(y_s|x_s = m)$ using finite mixtures of independent distributions:

$$p(y_s|x_s = m) = \sum_{i=1}^{K^{mnt}} \pi_i^{mnt} \, F_i^{mnt}(y_s|\theta_i^{mnt}), \qquad \forall s \in S_t^n \quad (21)$$

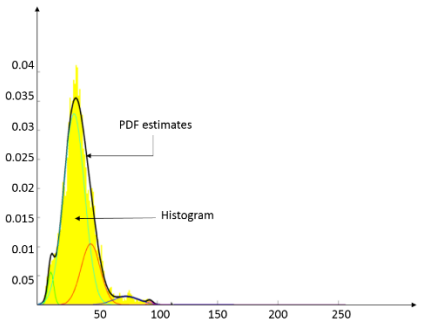


Figure 5: PDF modeling using finite mixtures

where π_i^{mnt} are the mixing proportions, θ_i^{mnt} is the set of the parameters of the i^{th} PDF mixture component of class m at scale level n and time t, and $F_i^{mnt}(\cdot)$ is the corresponding parametric family ($n = 0,1,\dots,R; m = 0,1,\dots,M-1; t = 0,1$).

When the data at scale level n and time t is an optical image, the class-conditional marginal PDF $p(y_s|x_s = m)$ related to each class m is modeled by a multivariate Gaussian mixture [34] with a set of

parameters associated with the corresponding means and covariance matrices. This Gaussian assumption, especially when combined with a finite mixture is a well-known and widely accepted model for the statistics of optical data. Thanks to the linearity of the wavelet operator, the same assumption also holds for the resulting transformed levels of the quad-tree.

The use of finite mixtures instead of single PDFs offers the possibility to consider heterogeneous PDFs, usually reflecting the contributions of different materials present in each class. This class heterogeneity is relevant when we address VHR images. The parameters of the mixture model θ_i^{mnt} in the Gaussian mixture are estimated through the stochastic expectation maximization (SEM) algorithm [31], which is an iterative stochastic parameter estimation algorithm developed for problems characterized by data incompleteness and approaching, under suitable assumptions, maximum likelihood estimates. For each scale and time, SEM is separately applied to the training simples of each class to estimate the related parameters. We note that SEM also automatically estimates the number of mixture components K^{mnt} [31]. Only the maximum number of such components has to be predefined (and, in our experiments, was fixed to 10).

III. EXPERIMENTAL RESULTS

1) Data Sets and Experimental Setup

In this section, we discuss the results of the experimental validation of the developed hierarchical classifier on two datasets (see figure 7):

- A three-date series of panchromatic and multispectral Pléiades images acquired over Port-au-Prince (Haiti) in 2011, 2012, and 2013.
- Two pan-sharpened GeoEye acquisitions acquired over Port-au-Prince (Haiti) in 2009 and 2010.

Five land cover classes have been considered for both data sets: urban (red), water (blue), vegetation (green), bare soil (yellow), and containers (purple). We note that these classes represent semantically high level land covers. However, a classification map associated with more detailed classes can be produced when a sophisticated ground truth is available. In the present work, manually annotated non-overlapping training and test sets were selected in homogeneous areas. Spatially disjoint training and test areas were used in all experiments to minimize correlation between training and test samples and prevent possible optimistic biases in accuracy assessment.

In the case of the Pléiades images, the finest resolution of the multiresolution pyramid (level 0) was set equal to the finest resolution of the input panchromatic images (i.e., 0.5 m). Co-registered multispectral images (at 2 m) were integrated in level 2 of the pyramid. To fill level 1, a wavelet decomposition of the panchromatic image was used. As a preliminary experiment, the combination of the proposed method with numerous wavelet operators, including Daubechies, biorthogonal, and reverse biorthogonal wavelets, symlets, and coiflets of various orders [33], was examined. The results were similar, and the main difference relied on the level of smoothness of the final classification map. On one hand, as shown in Figure 6, the average of the overall accuracies obtained on the test sets of all individual dates was remarkably stable as a function of the selection of the wavelet operator, suggesting that this selection is not critical in the application of the proposed approach. On the other hand, an exception was represented by the Daubechies wavelets of order 10 (db10) whose combination with the proposed multiresolution method resulted in higher accuracies than the other considered wavelet transforms. This wavelet operator will be used in all other experiments discussed in this paper (see Figure 6).

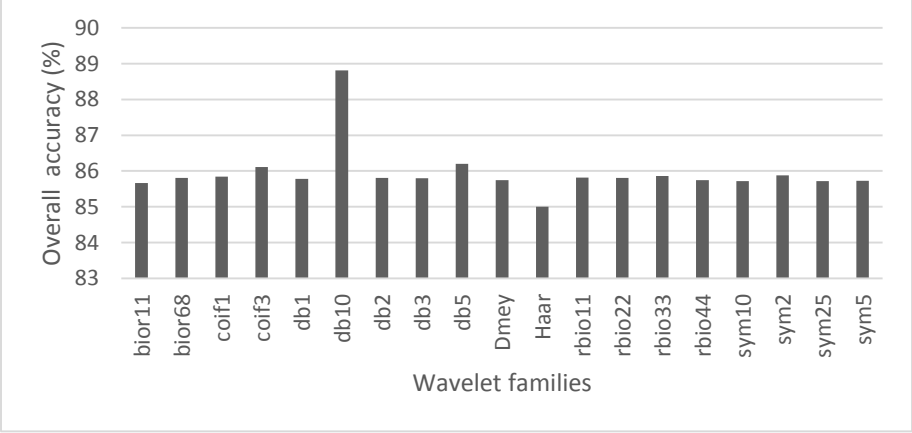


Figure 6: Pléiades data set: average of the overall accuracies obtained on the test sets of all individual dates results using several wavelet families.

The GeoEye image resolution is at 0.5 meters (the finest resolution of the multiresolution pyramid). Comments similar to those reported for the experiments with Pléiades images regarding the selection of the wavelet operator hold here as well, and db10 was used to fill level 1 and level 2 of the pyramid.

As discussed in Section II.A, the proposed method depends on four parameters, i.e., β in (10) and (15), θ in (12) and (20), φ in (20), and R. the classification results included in the paper were obtained using the following parameter values: $\beta = 0.8$, $\theta = 0.85$, $\varphi = 0.48$, and $R = 2$, i.e., three levels in each quad-tree. Including more decomposition levels generally assists in discriminating homogeneous land

covers but might result in the removal of small-size image details [1, 15, 51, 52]. Indeed, with $R = 2$, the classification map is generated at 50-cm spatial resolution for both data sets, while the coarsest scale corresponds to 2-m resolution. Including coarser-resolution features would generally favor spatial smoothness but may progressively hinder the capability to discriminate classes characterized by spatial details such as "buildings" and "containers."

The value of β was automatically optimized by applying the well-known pseudo-likelihood method to the training samples [27]. Accordingly, a user/operator does not have to perform a trial-and-error procedure to set β. In general, a lower value of β prevents spatial oversmoothing at the price of an accuracy decrease on test samples located inside homogeneous regions associated with the same thematic class.

As β is automatically optimized, the only free parameters are θ and φ. According to (20), θ is the probability that a site, its parent on the same date and its parent on the previous date all share the same class label; φ is the probability that a site shares the same label of one of the two parents while the parents disagree. For (20) to define a probability distribution, θ and φ can take values in [0, 1] and [0, 0.5], respectively, in the case of $M = 5$ classes. Figures 11 and 12 show the behavior of the overall accuracy of the proposed method on the test set as each one of these two parameters range in these intervals while the other parameter is fixed to the aforementioned reference values (i.e., $\theta = 0.85$ and $\varphi = 0.48$). On one hand, these plots suggest that the method is sensitive to the values of θ and φ. This is an expected result because they involve the causality of the model. On the other hand, limited sensitivity was observed and the overall accuracy remained higher than approximately 85% as long as φ and θ were larger than 0.4 and 0.85, respectively. On the contrary, it was basically for relatively extreme and not meaningful values of θ or φ that poor values of the overall accuracy were obtained. For example, $\varphi = 0.2$ implies that there is a 15% probability that the labels of a site and of its two parents are all different (see (20) with $M = 5$). Because of the very high interscale and temporal correlations associated with the multiresolution data sets, this event is highly unlikely, and $\varphi = 0.2$ yields to significantly overestimating its chances of occurring in the classification process, thus affecting the overall accuracy. Similarly, $\theta = 0.5$ implies a 12.5% probability that the parents of a site on the two dates agree on a certain class membership but the site disagrees, another outcome that is very unlikely because of interscale correlation and whose probability is overestimated.

Therefore, although the proposed method is overall sensitive to θ and φ, the experimental analysis suggests that the meaning of the two parameters in relation to interscale and temporal correlation allows

a user/operator to rather easily determine values of or ranges on these parameters that lead to accurate classification maps.

Preliminary experiments also pointed out that the use of the MMD optimization technique resulted in a significant reduction in the number of iterations needed to estimate the class label of each pixel as compared to the direct maximization of the posterior marginals. To reach convergence in the case of one single level of a quad-tree including 1600 x1000 pixels and 5 classes, MMD required fewer than 1000 iterations. We recall that, in each MMD iteration, one individual pixel is examined. On the contrary, the direct choice of the class label that maximizes the posterior marginal on each pixel required to iterate over all 1600 x1000 pixels and 5 classes, thus taking a much longer time.

The results obtained by the proposed method were compared to those generated by: the technique in [1], used as a multiresolution single-date benchmark in both its (i) MPM- and (ii) MAP-based formulations; (iii) the MRF-based algorithm in [2], used as single-resolution multitemporal benchmark; (iv) a contextual combination of the K nearest neighbor and MRF-based approaches, used as a single-resolution single-date benchmark; and (v) the well-known K-means clustering technique, used as a basic unsupervised benchmark. The results of the proposed and previous techniques are reported in the following subsection along with further details on their applications.

2) Experimental Results and comparisons

In this section, we present the classification maps and discuss the corresponding classification accuracies that were obtained on the test set. Figure 9 and Table 1 refer to the results obtained using Pléiades images, and Figure 10 and Table 2 regard those obtained from the GeoEye acquisitions All computation times reported in the tables refer to a C++ implementation on an Intel i7 quad-core (2.40 GHz) 8-GB-RAM 64-bit Linux system.

The analysis of the classification maps has suggested that the proposed hierarchical method leads to accurate results.

In particular, several experimental comparisons were performed with methods exploiting multi- or single-resolution, multi- or single-date, supervised or unsupervised approaches. First, the results of the proposed technique were compared to the separate hierarchical classifications results obtained at individual dates using the multiresolution single-time method in [1], in both its MPM (Figure 9(c) and Figure 10(c)) and MAP formulations and using a 3 level pyramid with the following parameters: $\beta = 0.8$

and $\theta = 0.85$ (see Figure 9(d) and Figure 10(d)). We recall that several extensions of the method in [1] have been developed including the approach presented by Voisin et al. in [15] for the specific case of multisensor classification and based on the integration of the hierarchical MRF model of Laferté et al. in [1] with copula functions for merging data from both optical and SAR sensors within the same pyramid. In the present paper, the focus is on multitemporal classification with optical images and not on multisensor fusion. Accordingly, we used the original method in [1] for comparison purposes. The results of the comparison suggest the effectiveness of the proposed multitemporal hierarchical model in fusing the temporal, spatial, and multiresolution information associated with the input data (see Table 1). In practice, the use of one quad-tree structure with the MPM criterion yields "blocky" segmentation (see Figure 8 (a)). This phenomenon can be explained by the fact that two neighboring sites at a given scale may not have the same parent. In this case, a boundary appears more easily than when they are linked by a parent node. These blocky artifacts are avoided by the use of the multitemporal hierarchical structure proposed in this work in which causal relationships between parents and offspring in the same quad-tree are relaxed by the introduction of other causal relationships over time and scale (see Figure 8(b)). One of the main sources of misclassification in the single-date results is the confusion between the "urban" and "vegetation" classes; this misclassification is reduced in the multitemporal classification obtained by the proposed method because of the modeling of the temporal relationships among the input multiresolution data. Furthermore, as expected, the MAP criterion was poorly effective when applied to the considered hierarchical structure because errors were propagated from the root to the leaves and led to severe misclassifications, especially regarding the classes that most strongly overlap in the feature space (e.g., "urban" and "containers"; see Figures 9 and 10 (d)).

Second, in the context of multitemporal classification, the proposed classifier was compared to the multitemporal single-resolution MRF-based method proposed in [2]. It uses the mutual approach and consists in performing a bidirectional exchange of the temporal information between the (non-hierarchical) single-time MRF models associated with consecutive images in the sequence. In the form of an appropriate energy function, each single-time MRF model integrates three types of information (spectral, spatial contextual, and temporal contextual) using a multilayer perceptron (MLP) neural network to extract the spectral information. The results reported in Table 2 show that a better exploitation of the spatio-temporal information allowed the proposed cascade multiresolution approach to provide more accurate results than the previous mutual single-resolution approach in [2]. More generally, the mutual approach reduces the risk of propagating the classification error between consecutive dates, while the use

of the hierarchical schema provided more accurate classification maps, at least, on the considered data sets. Furthermore, because of the hierarchical aspects and the non-iterative algorithm, only few minutes were necessary to obtain satisfactory results using the proposed approach compared to those obtained by the mutual approach that required a much longer computation time (several hours). According to the formulation of the method in [2], this time included the times required to compute the texture features from the given image time series, to train and apply an MLP neural network for the image of each date using the backpropagation algorithm, and to estimate the parameters of the corresponding MRF model using the case-specific parameter optimization procedure in [3].

The classification maps obtained using the well-known K-nearest-neighbors (K-NN) method are also shown in Figures 9 and 10 (h). K-NN was used as a benchmark non-parametric classifier. It is non-contextual, so to perform a fair comparison between the proposed method and a spatial-contextual technique, it was combined with an MRF model. A hidden MRF whose unary term was expressed in terms of the pixelwise posterior probabilities estimated by K-NN and whose contextual term was represented by an isotropic Potts model was used. K = 30 was estimated by cross validation on the training set, and the smoothing parameter of the Potts model was optimized using the automatic method in [36], which is based on the Ho-Kashyap algorithm. The numerical results on the test sets suggest that this single-scale MRF-based method (see Tables 1 and 2) leads to rather poor accuracy and severe spatial oversmoothing as shown in Figures 9 and 10(g). This is consistent with the fact that this combined K-NN + MRF classifier is intrinsically single-resolution and single-date, and can exploit neither the multiresolution nor the multitemporal structure of the input data set. In the map in Figure 10(g), obtained from GeoEye data, the combined K-NN + MRF well discriminated the “water” and “urban” classes but almost did not identify the other thematic classes due to the strong spectral overlapping and the imbalance between the training sample sizes of these classes.

Finally, a further comparison was performed between the results of the proposed method and those of an unsupervised algorithm. K-means was used for this benchmark comparison as a well-known consolidated approach, and was applied with K = 5. This number of clusters was used to match the number of classes in each data set. The clusters obtained by K-means generally do not coincide with the thematic classes of a supervised classification problem. An alternate strategy could be to, first, apply K-means using a significantly larger number of clusters, and then, perform a cluster-to-class assignment either manually or on the basis of the training set. In either case, this assignment would incorporate prior knowledge. This experiment was meant as a benchmark comparison with an unsupervised method using no prior

knowledge. Accordingly, the simple choice K = 5 was accepted. As expected due to its unsupervised, non-contextual, and single-resolution formulation, K-means performed the worst in terms of classification accuracy, while it exhibited the lowest computation time (see figures 9, 10 (h)).

Port au Prince
Pléiades ©CNES (2011), distribution Airbus DS

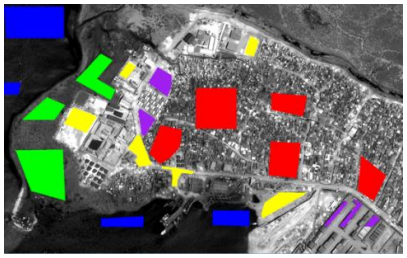
Port au Prince
Pléiades ©CNES (2012), distribution Airbus DS

Port au Prince
Pléiades ©CNES (2013), distribution Airbus DS

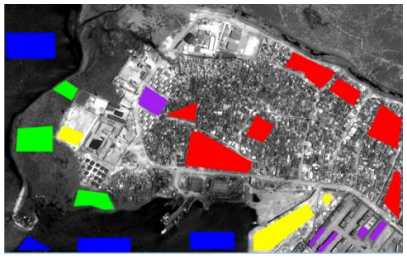
Port au Prince
©GeoEye (2009),

Port au Prince
©GeoEye (2010),

Figure 7: Examples of time series images acquired over Port-au-Prince (Haiti)



(a)

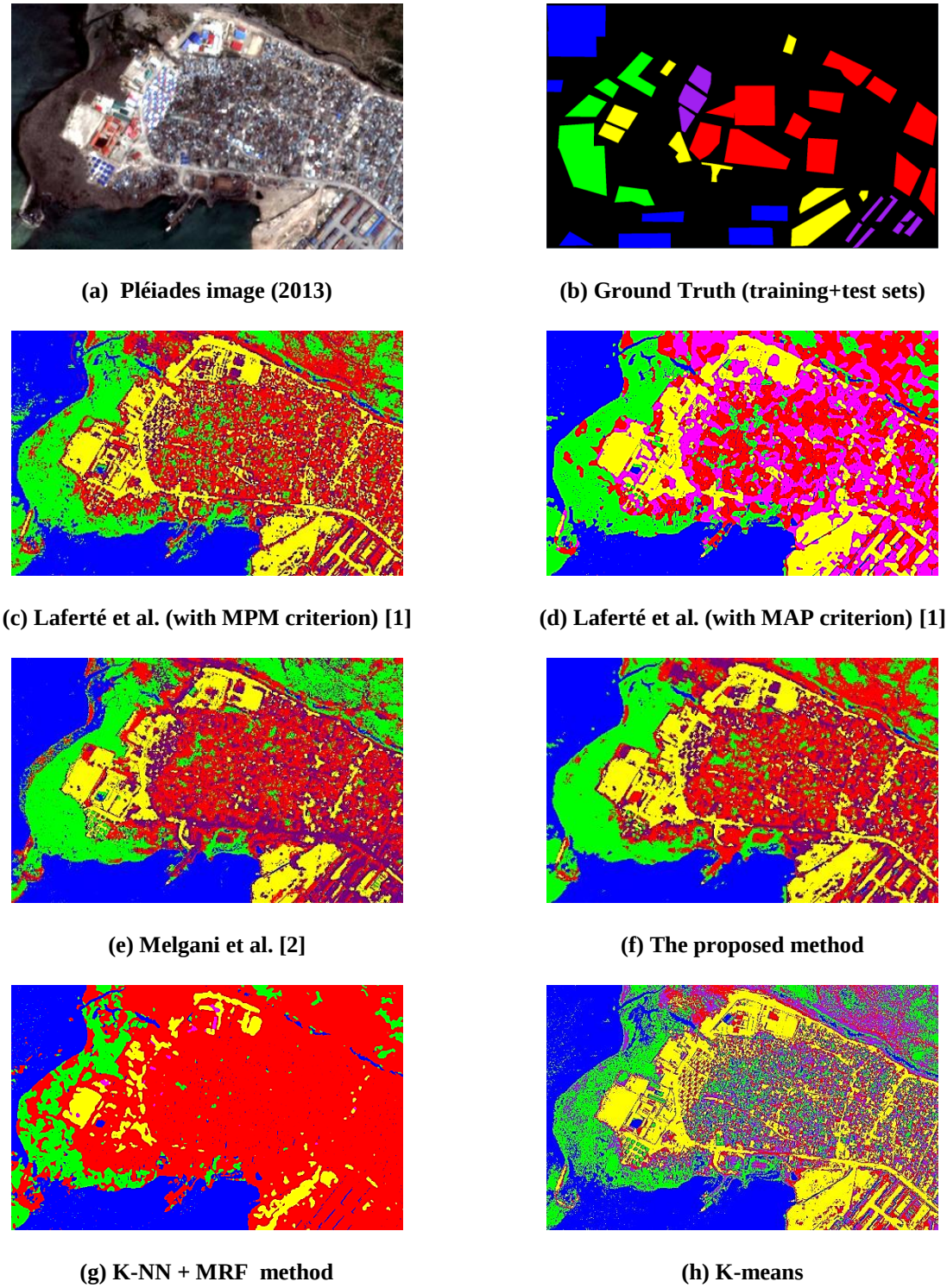


(b)

Figure 13: Ground truth for the Pléiades image acquired in 2013: (a) training set, (b) test set

Class name and color	# of pixels in training set	# of pixels in test set
Water(blue)	49 057	45 790
Urban (red)	75 508	72 327
Vegetation (green)	50 688	27 086
Bare soil (yellow)	29 333	25 541
Container (purple)	16 064	14 652

Table 3: Number of training and test samples on the panchromatic pixel lattice of the Pléiades image (1600 x1000 pixels) acquired in 2013



(a) Pléiades image (2013)

(b) Ground Truth (training+test sets)

(c) Laferté et al. (with MPM criterion) [1]

(d) Laferté et al. (with MAP criterion) [1]

(e) Melgani et al. [2]

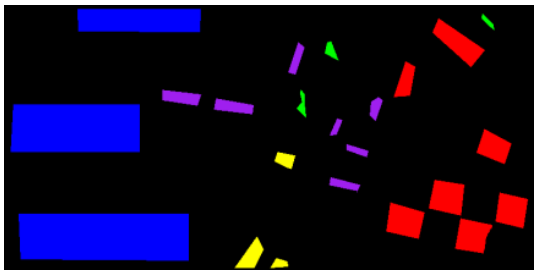
(f) The proposed method

(g) K-NN + MRF method

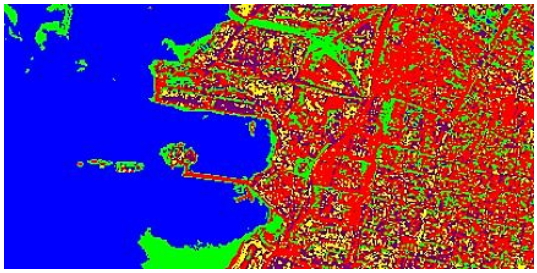
(h) K-means

Figure 9: Classification maps obtained from Pléiades images, (© CNES distribution Airbus DS). Legend: urban (red), water (blue), vegetation (green), bare soil (yellow) and containers (purple).

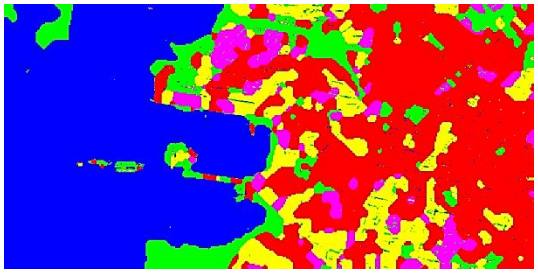
(a) GeoEye image (2010)



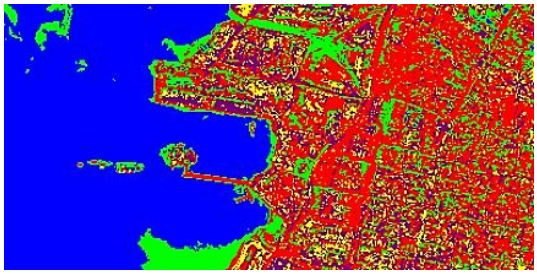
(b) Ground Truth (training+test sets)



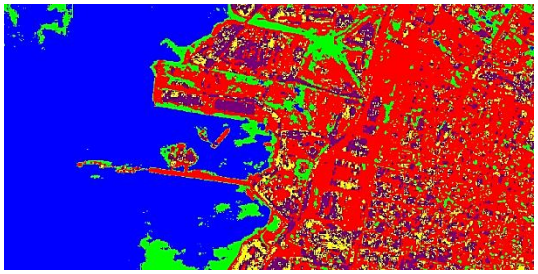
(c) Laferté et al. (with MPM criterion)



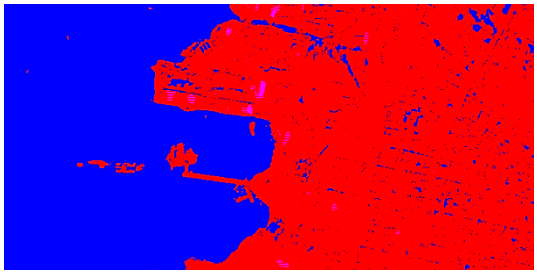
(d) Laferté et al. (with MAP criterion)



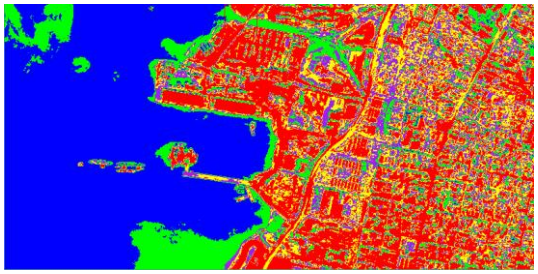
(e) Melgani et al.



(f) The Proposed method

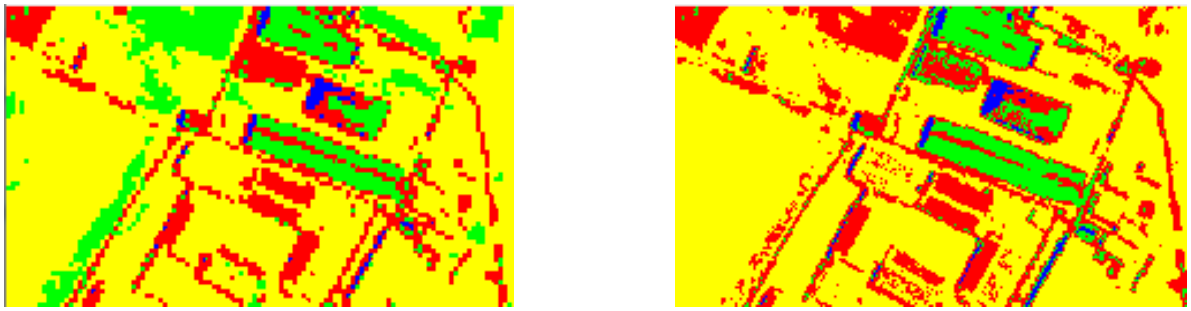


(g) KNN-MRF method



(h) K-means

Figure 10: Classification maps obtained from GeoEye images. Legend: urban (red), water (blue), vegetation (green), bare soil (yellow) and containers (purple).



(a) (b)

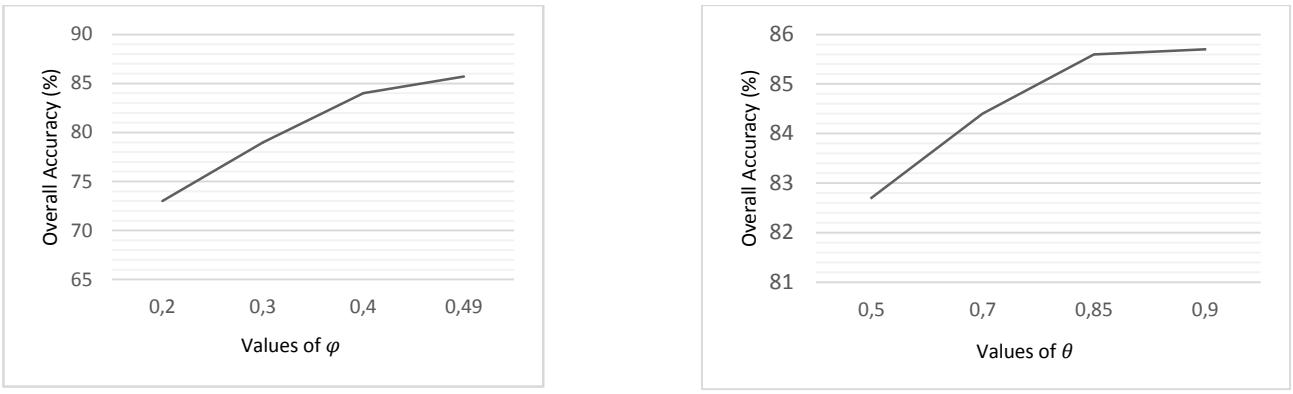
Figure 8: (a) Detail of two maps in Figure 7: blocky artefacts obtained using the method in [1] in its MPM formulation (b) reduction of these blocky artefacts using the proposed method.

	Port-au-Prince, Haiti						
	urban	**water**	**vegetation**	**bare soil**	**containers**	**overall accuracy**	**computation time**
Proposed method	81.62 %	100 %	90.69 %	92.82 %	62.82 %	**85,59 %**	**480 seconds**
Laferté et al. method using MPM criterion	77.45 %	88.62 %	72.59 %	86.02 %	57.02 %	**76.34 %**	**160 seconds**
Laferté et al. Method using MAP criterion	56.14 %	100%	81,90 %	87.02%	73.21%	**79.65 %**	**220 seconds**
Melgani et al. method	80.63 %	100 %	86.33 %	87.61 %	69.61 %	**84,83 %**	**≈1 hour**
K-NN + MRF	96,84%	92,42%	47.15 %	71.83 %	16.75 %	**64.99%**	**90 seconds**
K-means	12.37%	98.63%	59.18%	91.66 %	29.42 %	**58.25 %**	**20s**

Table 1. Classification accuracies on the test set of the Pléiades dataset: class accuracies (producer's accuracies), overall accuracy, and computation time.
Experiments were conducted using one (1600x1000) image in level 0, one (800x500) image in level 1 and four (400x250) bands in level 2 on an Intel i7 quad-core (2.40 GHz) 8-GB-RAM 64-bit Linux system.

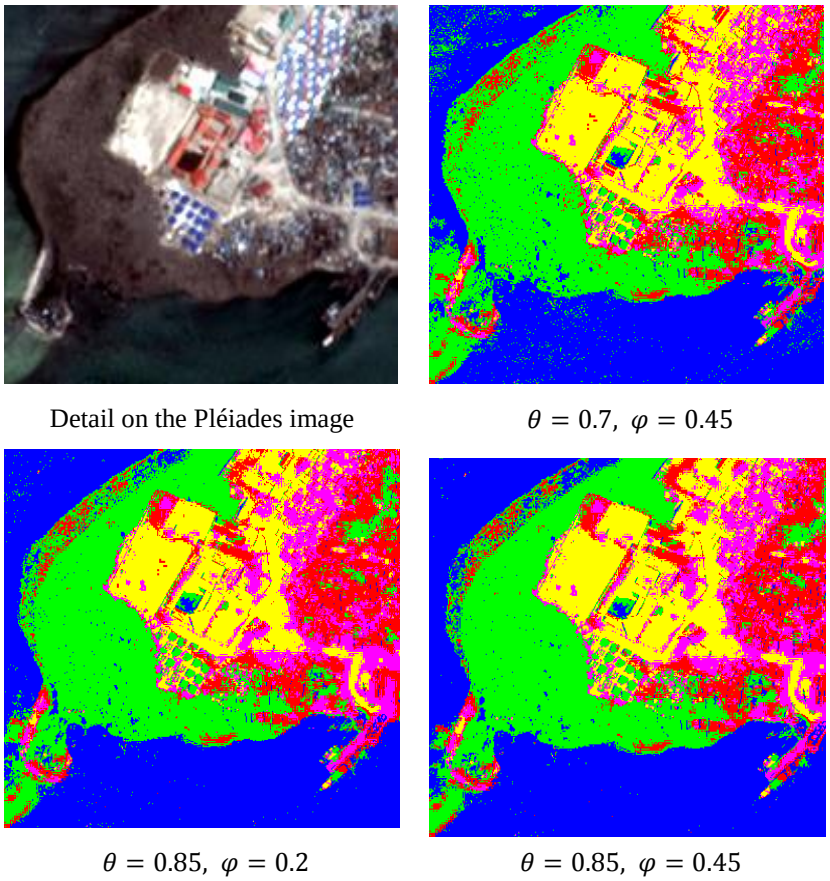
	Port-au-Prince, Haiti						
	urban	water	vegetation	bare soil	containers	**overall accuracy**	**computation time**
Proposed method	87.59 %	100 %	98.12 %	72.82 %	82.27 %	**88,16 %**	**345 seconds**
Laferté et al. method using MPM criterion	77.45 %	100 %	88.34 %	66.22 %	67.87 %	**79.97 %**	**90 seconds**
Laferté et al. method using MAP criterion	64.52 %	100%	92.15 %	85.62%	49.47 %	**78.35 %**	**140 seconds**
Melgani et al. method	80.63 %	100 %	89.79 %	70.54 %	74.29 %	**83,05 %**	**≈1 hour**
K-NN + MRF	100%	100%	0%	0%	12.28%	**42.45 %**	**40 seconds**
K-means	88.97%	100%	88.14%	45.6 %	36.96 %	**71.93 %**	**15 seconds**

Table 2. Classification accuracies on the test set of the GeoEye dataset: class accuracies (producer's accuracies), overall accuracy, and computation time.
Experiments were conducted using one (1600x800) image in level 0, one (800x400) image in level 1 and one (400x200) bands in level 2 on an Intel i7 quad-core (2.40 GHz) 8-GB-RAM 64-bit Linux system.



(a) (b)

Figure 11. Overall accuracy of the proposed method on the test set of the Pléiades data set as a function of the parameters (a) φ and (b) θ.



Detail on the Pléiades image $\theta = 0.7,\ \varphi = 0.45$

$\theta = 0.85,\ \varphi = 0.2$ $\theta = 0.85,\ \varphi = 0.45$

Figure 12. Details of the the classification maps obtained by the proposed method when applied to the Pléiades data set with different values of θ and φ.

IV. CONCLUSION

In the proposed method, multidate and multi-resolution fusion is based on explicit statistical modeling. The method combines a joint statistical model of the considered input optical images through hierarchical Markov random field modeling, leading to a statistical supervised classification approach. We have developed a novel MPM-based hierarchical Markov random field model that considers multitemporal information and, thus, supports the joint supervised classification of multiple images taken over the same area at different times and different spatial resolutions. We analyzed the results obtained with the proposed method through experiments with multitemporal Pléiades and GeoEye data sets. The experimental results suggest that the method is able to provide accurate classification maps. The proposed algorithm was compared to a previous single-date multiresolution method and a previous multidate single-resolution method, both based on MRF models associated with suitable (hierarchical or single-scale) pixel lattices and a couple of well-known classifiers including a contextual combination of the K nearest neighbor and MRF-based approaches, used as a single-resolution single-date benchmark; and the K-means clustering technique, used as a basic unsupervised benchmark. The proposed technique was demonstrated to be advantageous in terms of the classification accuracy on the test set, the spatial regularity of the classification maps, the minimization of spatial artifacts, and the tradeoff with respect to computation time. These results suggest the effectiveness of the algorithm in fusing both multitemporal and multiresolution information for supervised classification purposes and confirm that MRF models represent a powerful fusion tool in remote sensing.

The computational advantages of hierarchical MRFs, for which exact recursive formulations of the MPM decision rule are feasible with no need for time-expensive Metropolis or Gibbs sampling procedures, has also been confirmed by the experimental results of the proposed method and those of the benchmark single-resolution multiresolution classifier.

The proposed method is based on an MRF model on a case-specific topology that comprises multiple hierarchical quad-trees, each associated with an acquisition date. Wavelet transforms are used to fill in those levels of each quad-tree that are associated with input remote sensing imagery. The selection of the wavelet operator among a large family of possible transforms was not critical because most transforms lead to classification results with similar accuracies. Nevertheless, Daubechies wavelets of order 10 yielded higher accuracies than the other considered transforms. As a future extension of this method, the automation of the selection of the wavelet operator using, for example, a dictionary of multiple transforms [35] could be incorporated in the developed model and classifier. Similarly, the pseudo-likelihood method

was used to optimize the smoothing parameter of the Potts spatial component of the proposed MRF model. Alternate parameter estimation algorithms based, for example, on mean-square error [36], stochastic gradient, or Monte-Carlo techniques [37] could be integrated in the proposed method to address the optimization of this parameter and of the two parameters involved in the transition probabilities. The accuracy of the proposed method was found sensitive to these two parameters. However, the experimental results also suggested that, based on the meaning of these parameters in relation to temporal and interscale correlations, it is not difficult for a user/operator to identify ranges on their values that lead to meaningful models of the transition probabilities and yield accurate results.

In the proposed method, the number of classes is fixed for all levels of the hierarchical structure. When VHR data are considered, different types of land cover classes may be appreciated at different resolutions. Therefore, a further extension of this work would be to define different sets of classes at distinct levels of the pyramid and define a hierarchical link between these classes according to their semantic meaning. A pixel-level classification may not exceed the biophysical environment, but the neighborhood of the pixel brings substantial information that can be used to reconstruct landscape units and functional areas. Therefore, a semantic relationship between classes might be defined and would critically involve the availability of multiresolution ground truth data. Moreover, one main advantage of the proposed classifier is that it can be extended to be used for optical data, synthetic aperture radar (e.g., COSMO-SkyMed or RADARSAT-2) or multisensor data. The extension to the multisensor case will be a major direction of further research.

ACKNOWLEDGMENTS

The authors wish to thank the French Space Agency (Centre National des Etudes Spatiales, CNES) for providing the data used in the experiments and for partial financial support. The authors would also like to thank GeoEye Inc. and Google Crisis Response for providing the GeoEye imagery used for experiments.

ANNEX

Proof of equation (8)

With the same notations as in Section II, (8) is derived as follows:

$$
\begin{aligned}
p\left(x_s \mid y\right) &= \sum_{x_{s^-},\, x_{s^=}} p\left(x_s \mid y,\, x_{s^-}, x_{s^=}\right).\, p\left(x_{s^=}, x_{s^-} \mid y\right) \\
&= \sum_{x_{s^-},\, x_{s^=}} p\left(x_s \mid y_{d(s)},\, x_{s^-}, x_{s^=}\right).\, p\left(x_{s^=}, x_{s^-} \mid y\right) \\
&= \sum_{x_{s^-},\, x_{s^=}} \left[\frac{p\left(x_s\,,\, x_{s^-}\,, x_{s^=} \mid y_{d(s)}\right)}{\sum_{x_s} p\left(x_s\,,\, x_{s^-}\,, x_{s^=} \mid y_{d(s)}\right)}.\, p\left(x_{s^-}, x_{s^=} \mid y\right) \right] \\
&= \sum_{x_{s^-},\, x_{s^=}} \left[\frac{p\left(x_s\,,\, x_{s^-}\,, x_{s^=} \mid y_{d(s)}\right)}{\sum_{x_s} p\left(x_s\,,\, x_{s^-}\,, x_{s^=} \mid y_{d(s)}\right)}.\, p\left(x_{s^-} \mid x_{s^=}, y\right) p\left(x_{s^=} \mid y\right) \right] \\
&= \sum_{x_{s^-},\, x_{s^=}} \left[\frac{p\left(x_s\,,\, x_{s^-}\,, x_{s^=} \mid y_{d(s)}\right)}{\sum_{x_s} p\left(x_s\,,\, x_{s^-}\,, x_{s^=} \mid y_{d(s)}\right)}.\, p\left(x_{s^-} \mid y\right) p\left(x_{s^=} \mid y\right) \right]
\end{aligned}
$$

where the equalities across rows 1 and 2 and across rows 4 and 5 derive from assumptions A1 and A2, respectively (see Section II.B.2).

Proof of equation (9)

Again using the same notations as in Section II, (9) is obtained as follows:

$$
\begin{aligned}
p\left(x_s\,,\, x_{s^-}\,, x_{s^=} \mid y_{d(s)}\right) &= \; p\left(x_{s^-}\,, x_{s^=} \mid x_s,\, y_{d(s)}\right) \cdot\, p\left(x_s \mid y_{d(s)}\right) \\
&= \; p\left(x_{s^-}\,, x_{s^=} \mid x_s\right) \cdot\, p\left(x_s \mid y_{d(s)}\right) \\
&= \frac{p\left(x_s \mid x_{s^-}\,, x_{s^=}\right) \cdot\, p\left(x_{s^-}\,, x_{s^=}\right)}{p\left(x_s\right)} \cdot\, p\left(x_s \mid y_{d(s)}\right) \\
&= \; p\left(x_s \mid x_{s^-},\, x_{s^=}\right).\frac{p\left(x_{s^-} \mid x_{s^=}\right).\, p\left(x_{s^=}\right)}{p\left(x_s\right)} \cdot\, p\left(x_s \mid y_{d(s)}\right),
\end{aligned}
$$

where the equality across rows 1 and 2 derives from assumption A3 (see Section II.B.2).

REFERENCES

[1] J.M. Laferté, P. Perez and F. Heitz. "Discrete Markov modeling and inference on the quad-tree." IEEE Trans. Image Processing, 2000: 390–404.

[2] F. Melgani, S. B. Serpico. "A Markov Random Field Approach to Spatio-Temporal Contextual Image Classification." IEEE Transactions on Geoscience and Remote Sensing, vol. 41, No. 11, 2003: 2478-2487.

[3] Z. Kato and J. Zerubia. Markov Random Fields in Image Segmentation. Boston: NOW publishers, 2012.

[4] R. Chellappa and A. K. Jain, eds.,Markov Random Fields: Theory and Applications. Academic Press, 1993

[5] P. Fieguth, Statistical Image Processing and Multidimensional Modeling.New York, NY: Springer, 2011

[6] R. Kindermann and J. L. Snell,Markov Random Fields and their Applications.Providence, RI: American Mathematical Society, 1980

[7] S. Z. Li, Markov Random Field Modeling in Image Analysis. New York NY:Springer, 3rd Edition, 2009.

[8] Y. Rosanov, Markov Random Fields. New York, NY: Springer, 1982.

[9] G. Winkler, Image Analysis, Random Fields and Markov Chain Monte Carlo Methods. Springer, 2003

[10] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,"IEEE Trans. Pattern Anal. Machine Intel, vol. PAMI-6, pp. 721–741, 1984.

[11] S. B. Gelfand and S. K. Mitter, "On sampling methods and annealing algorithms," inMarkov Random Fields, (R. Chellappa, ed.), pp. 499–515, Boston, MA: Academic Press, 1993

[12] P. V. Laarhoven and E. Aarts,Simulated Annealing: Theory and Applications. Dordrecht: Kluwer Academic Publisher, 1987

[13] S. Rajasekaran, "On the convergence time of simulated annealing," Research Report MS-CIS-90-89, University of Pennsylvania, Department of Computer and Information Science, USA, November 1990.

[14] M. Luettgen, W. Karl and A. Willsky. "Efficient multiscale regularization with applications to the computation of optical flow." IEEE Trans. Image Processing, vol. 3, 1994: 41-64.

[15] A. Voisin, V. Krylov, G. Moser, J.Zerubia and S.B. Serpico. "Supervised Classification of Multi-sensor and Multi-resolution Remote Sensing Images with a Hierarchical Copula-based Approach." IEEE Trans. on Geoscience and Remote Sensing, 2014: 3346-3358.

[16] C. Bouman, "A multiscale image model for Bayesian image segmentation," Technical Report TR-EE 91-53, Purdue University, 1991.

[17] C. Bouman and B. Liu, "Multiple resolution segmentation of texture images,"IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, pp. 99–113, 1991

[18] P.H. Swain. "Bayesian classification in a time-varying environment." IEEE Trans. on Systems, Man and Cybernetics, vol 8, 1978.

[19] L. Bruzzone, D. Fernandez and S.B. Serpico. "A Neural-Statistical Approach to Multitemporal and Multisource Remote-Sensing Image Classification." IEEE Transactions on Geoscience and Remote Sensing, vol. 37, no. 3, 1999: 1350-1359.

[20] J. M. Jolion, A. Rosenfeld. A Pyramid Framework for Early Vision: Multiresolutional Computer Vision. Norwell, MA, USA: Kluwer Academic Publishers, 2004.

[21] M. Basseville, A. Benveniste and A. S. Willsky. "Multiscale autoregressive processes, Part I: Schur-Levinson parametrization." IEEE Trans. on Signal Processing, vol. 40. no. 8., 1992: 1915-1934.

[22] M. Basseville, A. Benveniste, K.C. Chou, S.A. Golden and R. Nikoukhah. "Modeling and Estimation of Multiresolution Stochastic Processes." IEEE Transactions on Information Theory, vol. 38, 2002: 766-784.

[23] T. S. Ferguson, Mathematical Statistics. A Decision Theoretic Approach. Probability and Mathematical Statistics. New York, NY: Academic Press, 1967.

[24] M. R. Luettgen, W. Karl, and A. S. Willsky, "Efficient multiscale regularization with applications to the computation of optical flow," IEEE Trans. Image Process., vol. 3, no. 1, pp. 41–64, 1994.

[25] G. Forney. "The Viterbi algorithm." Proc. IEEE, vol. 61, 1973: 268–278.

[26] L. Baum, T. Petrie, G. Soules and N.Weiss. "A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains." Ann. Math. Stat., vol. 41, 1970: 164–171.

[27] J. Besag, “Spatial interaction and the statistical analysis of lattice systems (with discussion),”J. R. Statist. Soc. B, vol. 36, pp. 192–326, 1974

[28] S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 6, no. 6, pp. 721–741, 1984.

[29] Z. Kato, J. Zerubia, and M. Berthod, “Satellite image classification using a modified Metropolis dynamics,” inProceedings of International Conference on Acoustics, Speech and Signal Processing, pp. 573–576, San Francisco, California, USA, March 1992.

[30] Z. Kato, J. Zerubia, and M. Berthod, “Bayesian image classification using Markov random fields,” in Maximum Entropy and Bayesian Methods, (A. Mohammad-Djafari and G. Demoment, eds.), pp. 375–382, Dordrecht Netherlands: Kluwer Academic Publisher, 1993.

[31] G. Celeux, D. Chauveau and J. Diebolt. "Stochastic versions of the EM algorithm: an experimental study in the mixture case." Journal of Statistical Computation and Simulation, vol. 55, no. 4, 1996: 287–314.

[32] C. Bouman, M. Shapiro. "A multiscale image model for Bayesian image segmentation." IEEE Trans. Image Processing, vol. 3, 1994: 162-177.

[33] S.Mallat. A Wavelet Tour of Signal Processing, 3rd ed. Academic Press, 2008.

[34] Mario A.T. Figueiredo., A. K. Jain, "Unsupervised learning of finite mixture models," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24, no.3, pp. 381-396, 2002.

[35] G. Moser, E. Angiati, S. B. Serpico, “Multiscale unsupervised change detection on optical images by Markov random fields and wavelets”, *IEEE Geosci. Remote Sensing Letters*, vol. 8, no. 4, pp. 725-729, 2011.

[36] S. B. Serpico, G. Moser, “Weight parameter optimization by the Ho-Kashyap algorithm in MRF models for supervised image classification”, *IEEE Trans. Geosci. Remote Sensing*, vol. 44, no. 12, pp. 3695-3705, 2006.

[37] M.V. Ibáñez and A. Simó, “Parameter estimation in Markov random field image modeling with imperfect observations. A comparative study,” *Pattern Recognition Letters*, vol. 24, pp. 2377–2389, 2003.

[38] Akbas, E.; Ahuja, N., "Low-Level Hierarchical Multiscale Segmentation Statistics of Natural Images," in Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.36, no.9, pp.1900-1906, Sept. 2014

[39] J. Deng, Y. Ban, J. Liu, L. Li, X. Niu and B. Zou, "Hierarchical Segmentation of Multitemporal RADARSAT-2 SAR Data Using Stationary Wavelet Transform and Algebraic Multigrid Method," in Geoscience and Remote Sensing, IEEE Transactions on , vol.52, no.7, pp.4353-4363,

[40] F. Petitjean, J. Inglada and P. Gancarski, “Satellite Image Time Series Analysis Under Time Warping,” IEEE Trans. on Geoscience and Remote Sensing, vol. 50, no. 8, pp. 3081-3095, 2012

[41] S. Martinis, A. Twele, S. Voigt, "Unsupervised Extraction of Flood-Induced Backscatter Changes in SAR Data Using Markov Image Modeling on Irregular Graphs," IEEE Transactions on Geoscience and Remote Sensing , vol.49, no.1, pp.251-263, 2011

[42] T. Hoberg, F. Rottensteiner, R.-Q. Feitosa and C. Heipke, "Conditional Random Fields for Multitemporal and Multiscale Classification of Optical Satellite Imagery," , IEEE Transactions on Geoscience and Remote Sensing , vol.53, no.2, pp.659-673, 2015

[43] Y. Boykov, O. Veksler, R. Zabih, "Fast approximate energy minimization via graph cuts," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.23, no.11, pp.1222-1239, 2001

[44] I. Hedhli, G. Moser, J. Zerubia, S.-B. Serpico. “New cascade model for hierarchical joint classification of multitemporal, multiresolution and multisensor remote sensing data”. IEEE ICIP – International Conference on Image Processing, Paris, France. 2014.

[45] A. Dawid “Applications of a general propagation algorithm for probabilidtic” expert systems,” Stat. Comput., vol. 2, pp. 25–36, 1992.

[46] F. Petitjean, J. Inglada and P. Gancarski, “Satellite Image Time Series Analysis Under Time Warping,” IEEE Trans. on Geoscience and Remote Sensing, vol. 50, no. 8, pp. 3081-3095, 2012

[47] F. Gao, T. Hilker, X. Zhu, M. Anderson, J. Masek, P. Wang, Y. Yang, "Fusing Landsat and MODIS Data for Vegetation Monitoring," IEEE Geosci. Remote Sensing Magazine, vol. 3, no. 3, pp. 47-60, 2015

[48] J.L. Marroquin, S. Mitter, T. Poggio. _ Probabilistic solution of ill posed problems in computational vision. J. American Statis. Assoc., 82:76_89, 1987.

[49] C. Bouman, M. Shapiro. "A multiscale image model for Bayesian image segmentation." IEEE Trans. Image Processing, vol. 3, 1994: 162-177.

[50] A. Singh, “Incremental estimation of image flow using a Kalman filter, in Proc. IEEE Workshop on Visual Motion, 1991, pp. 36-43.

[51] P. Pérez, A. Chardin, and J.M., Laferté, “Noniterative manipulation of discrete energy-based models for image analysis.” Pattern Recognition,33(4), pp.573-586, 2000.

[52] A. Chardin, and P. Pérez. "Semi-iterative inference with hierarchical models." In Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on, vol. 1, pp. 630-634. IEEE, 1998.